

---

# KruegerAlgorithms

Quantitative Marktdaten-Analyse

## Methodik der Algo-Entwicklung

Der ehrliche Workflow in fünf Phasen · ein Referenz-Durchlauf

*Kein neues Backtest-Ergebnis, sondern die Methodik dahinter: der fünfphasige Entwicklungs-Workflow von der Idee bis zum MT5-Algo — illustriert am vollständigen Cross-Asset-Durchlauf der Paper-Serie (DAX/FTSE/NQ/Dow, 2018–2026, netto Spread). Mit den Leitprinzipien, die echte von scheinbaren Edges trennen.*

### Dokument-Metadaten

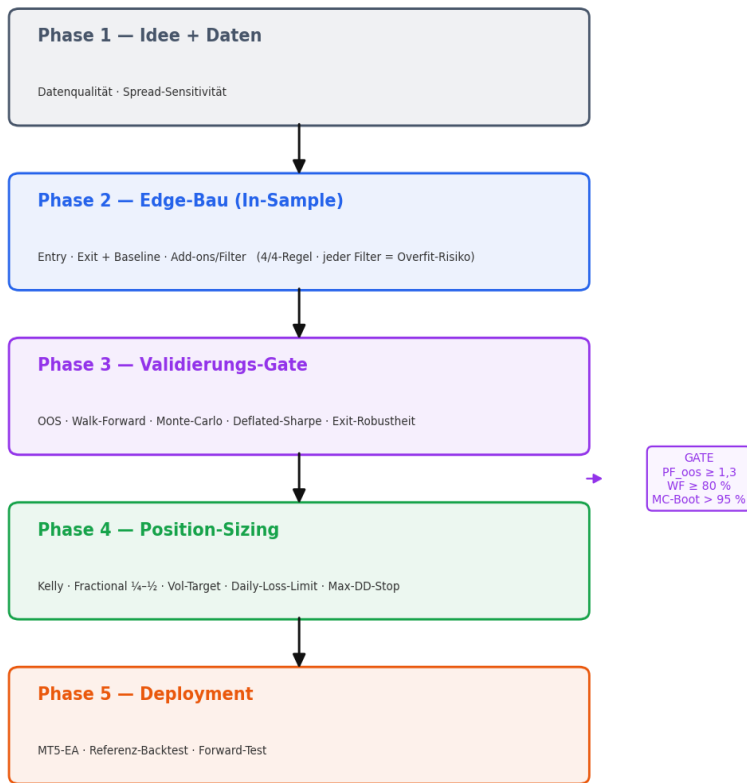
Feld	Wert
Herausgeber	KruegerAlgorithms
Titel	Methodik der Algo-Entwicklung
Erscheinungsdatum	6. Juni 2026
Version	v1.1
Typ	Methodik-Referenz (kein Backtest-Ergebnis)
Beispiel	Cross-Asset 2.-Kerze-Durchlauf, Phasen 1–5
Dokumenttyp	Methodische Darstellung (keine Trading-Empfehlung)

*Rein methodische Darstellung. Keine Trading-Empfehlung, keine Finanzberatung.*

## Worum es geht

Die meisten „profitablen Strategien“ sind überfittet: großartig auf einem Markt in einem Zeitraum, gescheitert live. Dieser Workflow soll das verhindern — durch eine Abfolge von Hürden, die jede Strategie überleben muss. Der Kern ist keine Technik, sondern eine Haltung: die Arbeit so aufbauen, dass sie sich selbst widerlegen kann.

### Der ehrliche Algo-Entwicklungs-Workflow



Querschnitt: netto Spread ab Phase 2 · Multiple-Testing-Korrektur · live Forward = einziges echtes OOS

Abb. 1: Der fünfphasige Workflow mit Gate und Trichter.

## Inhaltsverzeichnis

<b>Worum es geht</b>	<b>2</b>
<b>Inhaltsverzeichnis</b>	<b>3</b>
<b>1. Drei Leitprinzipien</b>	<b>4</b>
<b>2. Phase 1 — Idee &amp; Daten</b>	<b>4</b>
<b>3. Phase 2 — Edge-Bau (alles auf In-Sample)</b>	<b>4</b>
<b>4. Phase 3 — Das Validierungs-Gate</b>	<b>5</b>
<b>5. Phase 4 — Position-Sizing</b>	<b>5</b>
<b>6. Phase 5 — Deployment &amp; Realitätstest</b>	<b>6</b>
<b>7. Von Befund zu Algorithmus: wie man Merkmale einbaut</b>	<b>7</b>
7.1 Durchgerechnet: Conviction-Sizing auf „Short-Morgen-Stärke“ (OOS) . . . . .	7
<b>8. Der Referenz-Durchlauf &amp; Checkliste</b>	<b>8</b>
8.1 Checkliste / Entscheidungs-Gates . . . . .	8
8.2 Limitationen der Methodik . . . . .	8

## 1. Drei Leitprinzipien

**Der Trichter verengt sich.** Jede Phase wirft Kandidaten raus, nie hinein. Was in die Validierung getragen wird, ist ein Trial — jeder zusätzliche Trial erhöht die Zufalls-Bestehens-Chance. Wenige begründete Hypothesen schlagen viele optimierte.

**Kosten ab Phase 2, nicht am Ende.** Spread (und später Slippage) gehört in jede R-Zahl, sobald optimiert wird — nicht als Schönheitskorrektur zum Schluss. Was nur brutto funktioniert, funktioniert nicht.

**Live-Forward ist das einzige echte OOS.** Jeder historische OOS-Block stammt aus demselben Datensatz und ist nie ganz fremd. Der ehrliche Schluss-Test läuft erst ab heute.

## 2. Phase 1 — Idee & Daten

Die Idee mechanisch fassen (Beispiel: Ausbruch aus der 2. 15-Min-Kerze nach Cash-Open) und die Daten ehrlich darstellen, bevor optimiert wird: Quelle, Zeitraum, Lücken, und eine Spread-Sensitivität — wie viel des Brutto-Edges überlebt realistische Kosten? Wer hier overclaimt, baut auf Sand.

## 3. Phase 2 — Edge-Bau (alles auf In-Sample)

Reihenfolge: (1) Entry-Logik. (2) Exit-Logik + Baseline — erst eine Baseline definieren, dann jede Modifikation dagegen messen. (3) Add-ons/Filter — Pre-Trade-Filter UND In-Trade-Add-ons (Pyramiding). Genau hier, nicht später.

Exit-Variante (Mittel über 4 Indizes)	Mean-PF	profitabel
Time-BE 30 min	1,52	4/4
Time-BE 60 min	1,30	4/4
Exit@ONR-Kante + BE60	1,23	4/4
Baseline (Halten bis SL/EOD)	1,09	3/4
Trailing / Teil-TP / Multi-TP	$\leq 1,09$	0–3/4

Tabelle 1: Nur Zeit-Break-Even schlägt die Baseline auf allen vier Indizes; aktive Stop-Verwaltung kostet Erwartungswert.

*Jeder Filter, jedes Add-on ist ein zusätzlicher Parameter — und damit ein Overfit-Risiko. Hürde: Wirkung auf ALLEN Märkten (4/4-Regel), sonst Curve-Fit. Im Beispiel generalisierte von sieben Filtern nur der Short-Bias; Pyramiding senkte in jeder Größe das Kelly-Wachstum (optimale Add-Größe = 0). Beides ehrlich verworfen. Add-ons bewertet man nach dem Kelly-Kriterium (geometrisches Wachstum) — es bestraft Varianz und Drawdown automatisch.*

### 4. Phase 3 — Das Validierungs-Gate

Das Regelwerk ist ab hier fix. Vier unabhängige Verfahren prüfen, ob der Edge auf nicht-selektierten Daten überlebt: Out-of-Sample; Walk-Forward (rollierend + anchored); Monte-Carlo (Shuffle für Drawdown + Bootstrap für echte Erfolgswahrscheinlichkeit); Deflated-Sharpe (Multiple-Testing-Korrektur); Exit-Robustheit. **Das strenge Gate:**  $PF_{\text{oos}} \geq 1,3$  UND  $\geq 80\%$  profitable WF-Fenster UND  $Recovery \geq 2$  UND  $Bootstrap-P > 95\%$ .

Index	IS-PF	OOS-PF	WF-prof	Bootstrap-P	GATE
NQ	1,50	1,52	100 %	100 %	BESTANDEN
DAX	1,50	1,27	76 %	99,9 %	knapp durch
DOW	1,51	1,16	91 %	100 %	durch (PF)
FTSE	1,50	0,84	49 %	56,5 %	klar durch

Tabelle 2: In-Sample sahen alle vier identisch aus (PF ~1,50) — nur NQ besteht das Gate. Genau das ist der Sinn: der In-Sample-PF lügt.

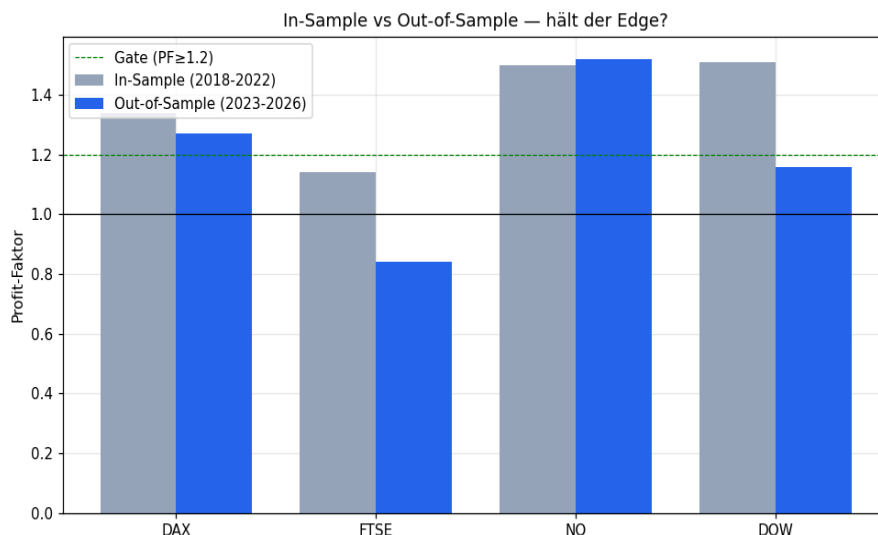


Abb. 2: In-Sample- vs Out-of-Sample-PF — die Spreu trennt sich erst out-of-sample.

### 5. Phase 4 — Position-Sizing

Erst nach bestandenem Gate — und nur auf dem Kern, der bestand (NQ). Kelly (optimal-f via  $\text{Max } E[\log(1+f \cdot r)]$ ), Fractional Kelly ( $\frac{1}{4}$ – $\frac{1}{2}$ , da Voll-Kelly praktisch Selbstmord), Vol-Targeting und harte Risiko-Overlays (Daily-Loss-Limit, Max-DD-Stop).

Sizing (NQ, OOS)	Risiko/Trade	CAGR	MaxDD
Voll-Kelly	~10 %	extrem	74–89 % (Selbstmord)
Halb-Kelly	~5 %	hoch	~45 %
Viertel-Kelly (praktikabel)	~2,5 %	+170 %*	~27 %

Tabelle 3: Fractional Kelly als praktikable Untergrenze. \*Idealisiert; real deutlich weniger — siehe Phase 5.

## 6. Phase 5 — Deployment & Realitätstest

Das fertige System als lauffähiger MQL5-EA, plus der ECHTE MT5-Lauf gegen den ehrlichsten Maßstab: Buy-and-Hold des Index. Erst hier zeigt sich, was der idealisierte Backtest wert ist.

NQ, 2021–2026, 1%/Trade	Total	MaxDD
Python-Referenz (idealisiert)	+963 %	11,7 %
MT5 REAL (Pepperstone)	+273 %	34,5 %
Buy-and-Hold (Index halten)	+127 %	35,6 %

Tabelle 4: NQ schlägt Buy-and-Hold real  $\sim 2,1\times$  — aber reale Kosten kosten  $-72\%$  Rendite und verdreifachen den Drawdown gegenüber dem Backtest. (DAX real  $-1\%$   $\rightarrow$  tot.)

**Edge-Vergänglichkeit — die unbequemste Lehre.** Der echte NQ-Lauf war stark bis Ende 2024, danach negativ (2025  $-9\%$ , 2026  $-11\%$ ) — während der Backtest grün blieb und es NICHT an Kosten lag. Ein Backtest kann lügen, selbst ein sauberer. Nur der Live-/Forward-Test sagt die Wahrheit; mechanischer Betrieb braucht einen rollierenden PF-Kill-Switch.

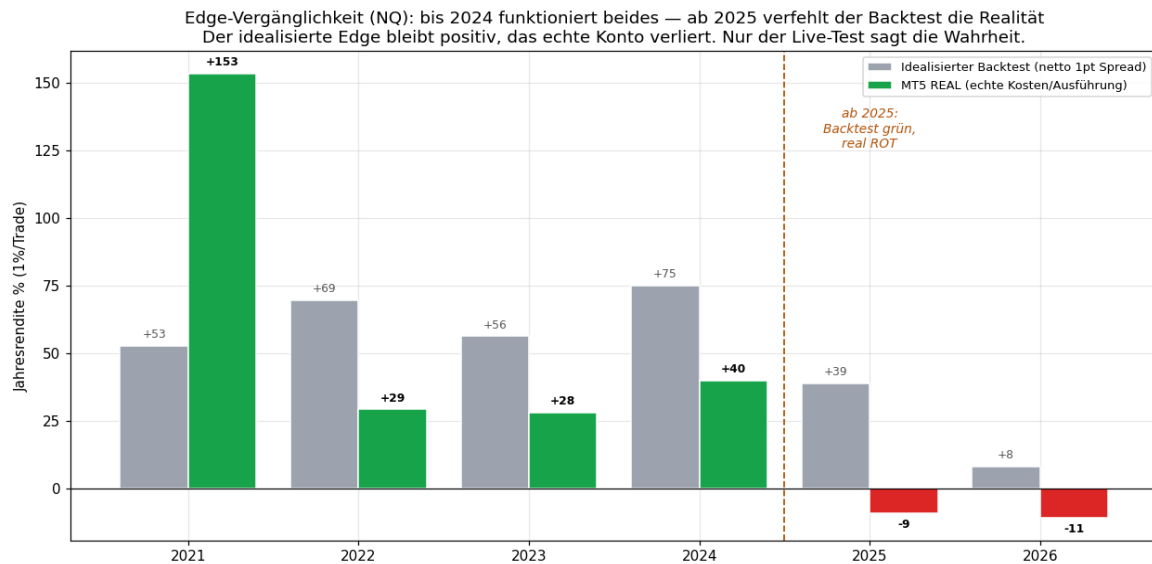


Abb. 3: Edge-Vergänglichkeit — ab 2025 verfehlt der Backtest die Realität (grün vs rot).

## 7. Von Befund zu Algorithmus: wie man Merkmale einbaut

Ein gefundenes Merkmal (z. B. „Gap-up → Short besser“) ist noch keine Regel. Vier übliche Wege, einen Befund in einen Algorithmus zu gießen — von überfit-anfällig bis robust:

Weg	Prinzip	Risiko
1) Harter Filter	Trade nur bei erfüllter Bedingung	hoch (Schwelle = Parameter, Klippen-Effekt)
2) Richtungs-Bias	Merkmal wählt die Seite, nicht das Ob	mittel
3) Conviction-Sizing	Größe skaliert mit erwartetem Edge	niedrig (kein Cutoff)
4) Score/Modell	mehrere Merkmale → ein Score	sehr hoch

Tabelle 5: Vier Wege, einen Merkmals-Befund zu nutzen.

*Faustregel: Weg 3 (Risiko-/Größen-Anpassung) ist häufig die robusteste Variante. Sie braucht kein hartes Cutoff, nutzt ein schwaches Signal graduell statt sprunghaft und hängt nicht von einer einzelnen gefitteten Schwelle ab. Harte Filter (Weg 1) sehen im Backtest oft besser aus, sind aber am anfälligsten für genau die Selbsttäuschung, die diese Methodik verhindern soll.*

**Disziplin (für alle vier gleich):** Selektion ≠ Validierung (separates Set, durchs Gate); Trial-Count im Deflated-Sharpe; wenige, cross-asset-konsistente Merkmale (4/4); kontinuierlich schlägt binär; korrelierte Merkmale sind kein additiver Edge (z. B. „Schluss über ONR“ und „Gap-up“ messen beide Morgen-Stärke — stapeln verdoppelt den Vorteil nicht).

### 7.1 Durchgerechnet: Conviction-Sizing auf „Short-Morgen-Stärke“ (OOS)

Konkreter Test von Weg 3: drei Tiers (strong = Short + Morgen-Stärke; avoid = Long gegen Morgen-Schwäche; neutral). Gewichte NUR aus In-Sample ( $\leq 2022$ ), mittel-normiert (gleiche Durchschnitts-Exposure → kein versteckter Hebel), out-of-sample (2023–2026) angewendet. Gepoolt über vier Indizes:

OOS-Sizing	Total-R	meanR	Sharpe
Flat (1×)	253	0,075	1,39
Conviction	289	0,086	1,36

Tabelle 6: Conviction-Sizing vs Flat, out-of-sample, mittel-normiert.

*Befund: Das Sizing verteilt korrekt Richtung der höher-erwarteten Shorts um → +14 % Total-R, aber kein Sharpe-Gewinn (1,39 → 1,36) — die Mehrrendite kam mit proportional mehr Varianz (pro Index gemischt). Die Lehre: Conviction-Sizing ist der richtige, klippenfreie Weg, ein Signal zu nutzen — aber es zaubert keinen Edge herbei, der nicht robust da ist. Hätte man nur In-Sample geschaut, hätte man die +14 % als „Erfolg“ verbucht. Disziplin schlägt Technik.*

## 8. Der Referenz-Durchlauf & Checkliste

Was die fünf Phasen aus vier Kandidaten gemacht haben:

Markt	In-Sample-PF	Phase-3-Gate	Phase-5-Rolle
NQ	~1,50	bestanden	Kern-Algo
DAX	~1,50	grenzwertig	sekundär, etikettiert
DOW	~1,50	durch (PF)	verworfen
FTSE	~1,50	klar durch	verworfen

Tabelle 7: Vier in-sample identische Märkte — drei aussortiert. Das ist der Wert der Methodik, nicht eine einzelne Kennzahl.

### 8.1 Checkliste / Entscheidungs-Gates

- Phase 1: Idee mechanisch? Daten ehrlich? Spread-Sensitivität gerechnet?
- Phase 2: Baseline definiert? Jede Modifikation dagegen gemessen? Add-ons auf 4/4, netto Spread?
- Phase 3: Regelwerk fix? OOS + WF + MC + DSR + Robustheit? Gate-Kriterien hart VOR dem Ergebnis definiert?
- Phase 4: Nur auf dem gate-konformen Kern? Fractional statt Voll-Kelly? Risiko-Overlays?
- Phase 5: Code = Backtest-Logik 1:1? Forward-Test angesetzt?

### 8.2 Limitationen der Methodik

*Auch dieser Workflow ist kein Garant. Der OOS-Block stammt aus demselben Datensatz; die Gate-Schwellen (1,3 / 80 % / 95 %) sind begründete, aber gesetzte Werte; der Deflated-Sharpe ist annahmesensitiv; die endgültige Wahrheit liefert nur der Live-Betrieb. Die Methodik minimiert das Risiko der Selbsttäuschung — sie eliminiert es nicht.*

*Erstellt und veröffentlicht von KruegerAlgorithms. Datenstand 6. Juni 2026. Version v1.1. Capstone der Paper-Serie zur 2. Eröffnungskerze.*